

# BigARTM: Open Source Library for Regularized Topic Modeling of Large Text Collections

Konstantin Vorontsov<sup>1,3</sup>, Oleksandr Frei<sup>2</sup>, Murat Apishev<sup>3</sup>, and Peter Romov<sup>4</sup>

<sup>1</sup> Department of Intelligent Systems at Dorodnicyn Computing Centre of RAS,  
Moscow Institute of Physics and Technology, [voron@forecsys.ru](mailto:voron@forecsys.ru)

<sup>2</sup> Schlumberger ...

<sup>3</sup> Moscow State University, ...

<sup>4</sup> Yandex ...

**Abstract.** BigARTM

**Keywords:** probabilistic topic model, Probabilistic Latent Semantic Analysis, Latent Dirichlet Allocation, stochastic matrix factorization, regularization, EM-algorithm, BigARTM.

## 1 Introduction

[?] [?] [?]  
[?] [?] [?]

## 2 Multimodal regularized topic model

multimodal LDA (mmLDA) and correspondence LDA (Corr-LDA) [?]  
Matching Words and Pictures, MoM-LDA [?]  
[?]  
[?]  
[?]

## 3 Online topic modeling

[?]  
[?]

## 4 BigARTM parallel distributed architecture

## 5 Experiments

## 6 Conclusions

**Acknowledgements.** The work was supported by the Russian Foundation for Basic Research grants 14-07-00847, 14-07-00908, and by Skolkovo Institute of Science and Technology (project 081-R).